BEST AVAILABLE COPY

PCT/GB 2004 / 0 0 4 9 2 8

The Patent Office

PATENTS • DESIGNS • TRADE MARKS • COPYRIGHT

REC'D 07 JAN 2005

WIPO PCT

INVESTOR IN PEOPLE

The Patent Office
Concept House
Cardiff Road
Newport
South Wales
NP10 8QQ

PRIORITY DOCUMENT

SUBMITTED OR TRANSMITTED IN COMPLIANCE WITH RULE 17.1(a) OR (b)

I, the undersigned, being an officer duly authorised in accordance with Section 74(1) and (4) of the Deregulation & Contracting Out Act 1994, to sign and issue certificates on behalf of the Comptroller-General, hereby certify that annexed hereto is a true copy of the documents as originally filed in connection with the patent application identified therein.

In accordance with the Patents (Companies Re-registration) Rules 1982, if a company named in this certificate and any accompanying documents has re-registered under the Companies Act 1980 with the same name as that with which it was registered immediately before re-registration save for the substitution as, or inclusion as, the last part of the name of the words "public limited company" or their equivalents in Welsh, references to the name of the company in this certificate and any accompanying documents shall be treated as references to the name with which it is so re-registered.

In accordance with the rules, the words "public limited company" may be replaced by p.l.c., plc, P.L.C. or PLC.

Re-registration under the Companies Act does not constitute a new legal entity but merely subjects the company to certain additional company law rules.

Signed Stephen Hordley

Dated 22 December 2004

Patents Form 1/77

Patents Act 1977
(Rule 16)

The Patent Office

25NOV03 E854705-4 D02884
P01/7700 0.00-0327338.0

THE PATENT OFFICE
C
25 NOV 2003
NEWPORT

The Patent Office

Cardiff Road
Newport
South Wales
NP10 8QQ

Request for grant of a patent

(See the notes on the back of this form. You can also get an explanatory leaflet from the Patent Office to help you fill in this form)

1. Your reference — P339851/CMU/RTH/RMC

2. Patent application number *(The Patent Office will fill in this part)* — 0327338.0 — 25 NOV 2003

3. Full name, address and postcode of the or of each applicant *(underline all surnames)* — The Queen's University of Belfast, University Road, Belfast, BT7 1NN

   Patents ADP number *(if you know it)* — 889659006

   If the applicant is a corporate body, give the country/state of its incorporation — United Kingdom

4. Title of the invention — "Assay"

5. Name of your agent *(if you have one)* — Murgitroyd & Company

   "Address for service" in the United Kingdom to which all correspondence should be sent *(including the postcode)* — Scotland House, 165-169 Scotland Street, Glasgow, G5 8PL

   Patents ADP number *(if you know it)* — 1198015

6. If you are declaring priority from one or more earlier patent applications, give the country and the date of filing of the or of each of these earlier applications and *(if you know it)* the or each application number

Country	Priority application number *(if you know it)*	Date of filing *(day / month / year)*

7. If this application is divided or otherwise derived from an earlier UK application, give the number and the filing date of the earlier application

Number of earlier application	Date of filing *(day / month / year)*

8. Is a statement of inventorship and of right to grant of a patent required in support of this request? *(Answer 'Yes' if:*
   a) any applicant named in part 3 is not an inventor, or
   b) there is an inventor who is not named as an applicant, or
   c) any named applicant is a corporate body.
   See note (d)) — Yes

9. Enter the number of sheets for any of the following items you are filing with this form. Do not count copies of the same document

Continuation sheets of this form	-----
Description	47
Claim *(s)*	10
Abstract	-----
Drawing *(s)*	4 + 4 RM.

10. If you are also filing any of the following, state how many against each item.

Priority documents	
Translations of priority documents	
Statement of inventorship and right to grant of a patent *(Patents Form 7/77)*	
Request for preliminary examination and search *(Patents Form 9/77)*	
Request for substantive examination *(Patents Form 10/77)*	
Any other documents *(please specify)*	

11. I/We request the grant of a patent on the basis of this application.

Signature Murgitroyd & Company

Date 24 November 2003

12. Name and daytime telephone number of person to contact in the United Kingdom

MALCOLM MAIN 0141 307 8400

Warning

After an application for a patent has been filed, the Comptroller of the Patent Office will consider whether publication or communication of the invention should be prohibited or restricted under Section 22 of the Patents Act 1977. You will be informed if it is necessary to prohibit or restrict your invention in this way. Furthermore, if you live in the United Kingdom, Section 23 of the Patents Act 1977 stops you from applying for a patent abroad without first getting written permission from the Patent Office unless an application has been filed at least 6 weeks beforehand in the United Kingdom for a patent for the same invention and either no direction prohibiting publication or communication has been given, or any such direction has been revoked.

Notes

a) If you need help to fill in this form or you have any questions, please contact the Patent Office on 08459 500505.

b) Write your answers in capital letters using black ink or you may type them.

c) If there is not enough space for all the relevant details on any part of this form, please continue on a separate sheet of paper and write "see continuation sheet" in the relevant part(s). Any continuation sheet should be attached to this form.

d) If you have answered 'Yes' Patents Form 7/77 will need to be filed.

e) Once you have filled in the form you must remember to sign and date it.

f) For details of the fee and ways to pay please contact the Patent Office.

"Detection of Protein Interactions"

Field of the Invention

The present invention relates to a method of detecting interactions between macromolecules. In particular, but not exclusively, the invention relates to a method of detecting protein interactions using fluorescence.

Background to the Invention

Protein to protein interactions play a key role in many biological processes including the assembly of enzymes, protein homo/hetero-oligomers, regulation of intracellular transport, gene expression, receptor-ligand interactions, entry of pathogens into the cell and the action of small molecules or drugs.

Identification and characterisation of macromolecular interactions can be performed using co-immunoprecipitation from cell lysates and solubilised membranes. However, this technique requires specific antibodies for both capture and identification of proteins and may further require the use of detergent to disrupt interactions.

More recently non-invasive techniques have been developed to determine protein to protein interactions.

Such non-invasive techniques were pioneered by the yeast two hybrid method which is based on complementation of a split yeast nuclear transcription factor. The yeast two hybrid method utilises a mammalian bait protein connected to a yeast DNA-binding domain. This bait protein is used to determine which prey proteins are able to bind to the bait protein from a mixture of prey proteins. The mammalian prey protein is connected to a yeast transcription activation domain. When the mammalian bait and prey proteins interact, the yeast DNA-binding and transcription activating domains are brought together. The DNA binding domain can bind to the yeast DNA and the transcription activating domain is then suitably located to trigger the expression of a reporter gene encoding an enzyme which in turn can catalyse the production of a coloured product within the yeast cells thus

indicating a successful interaction of bait with prey.

The use of yeast expression systems to identify mammalian protein-to-protein interaction suffers from a number of flaws. Certain post-translational modifications, that are normally critical to mammalian protein interactions, cannot be achieved by yeast cells. For example, tyrosine phosphorylation is key to many mammalian intracellular protein binding events involved in signal transduction. However, the yeast genome contains no tyrosine kinase genes so phosphotyrosine-dependent protein interactions cannot be accessed in yeast two hybrid studies.

Furthermore, in yeast two hybrid screening the protein complex must be able to translocate to the nucleus to cause expression of the reporter gene or cause downstream events to trigger the expression of a reporter gene. Thus proteins that are excluded from the yeast nucleus will not be accessible to this screening method.

Further methods such as protein complementation and the split ubiquitin method utilise similar underlying concepts to the yeast two hybrid method in that the interaction of two proteins (a bait and prey protein) act to express a reporter protein, the reporter gene allowing the interaction event to be visualised as a detectable signal.

Such methods which utilise the expression of a reporter enzyme to produce a detectable signal suffer from the disadvantage that the location of the protein complexes being detected cannot be accurately visualised in the cell.

Recently the technique of fluorescence energy transfer (FRET) has been used to determine protein to protein interactions. In this technique the interaction of two fluorophores indicates their close spatial proximity. For protein to protein interaction monitoring the addition of an absorbing moiety to one protein partner is complemented by the addition of a second fluorescing moiety to the second binding partner. Provided the emission spectrum of the absorbing moiety overlaps the excitation spectrum of the fluorescing moiety and both moieties are within 100Å of each other FRET will occur. Mutations in the sequence of green fluorescent protein (GFP) from the jellyfish *Aequorea victoria* have been studied and shown to cause variations in the spectral emission of GFP giving rise to variants of GFP such as Yellow Fluorescent Protein (YFP), as well as cyan (CFP) and blue (BFP) fluorescing variants. This technique uses fluorescent energy transfer between these colour variants of GFP which are fused to interacting proteins to determine protein to protein interaction. Using this method, when the two GFP derived fluorophores are brought into close proximity, energy transfer between the fluorescent variants occurs and changes in fluorescence

emissions are detected. Unfortunately, this method requires overexpression of the GFP fusion proteins to allow quantification of the small changes in fluorescence. Related methods to FRET require the use of irreversible photobleaching (FRAP) or expensive instruments capable of measuring fluorescence lifetime imaging (FLIM).

As a preliminary to the current experiments it was shown that green fluorescent protein can be engineered to add amino acid residues at particular regions in the GFP sequence whilst fluorescence is retained. Further, it has been shown in Hu, CD, Chinenov, Y.. and Kerppola, T. K. (2002). *Mol. Cell.* 9, 789-798 that using recombinant DNA technology specific Yellow Fluorescent Protein (YFP) fragments covalently fused to peptide sequences, which are capable of interacting with each other can reconstitute a fluorophore when the YFP fragments are brought together, such that the peptide sequences could interact.

Further, it has recently been shown that fluorescence can be generated following the functional association of two separate fragments of the GFP molecule (hapto-GFPs) when driven by the interaction of a pair of proteins fused via a linker to the new C′ and N′ termini of the hapto-GFPs. (Ghosh et al, (2000); Hu et al, (2002).

However, the above methods suffer from the disadvantage that functional association of

fluorescent fragments is limited by the constraints of stereochemistry imposed on the fragments by the bait and prey proteins' association. If the fusion termini of the interacting partners are widely separated, productive association of the haptoGFPs will not occur and no signal will be generated to indicate the interaction between the bait and prey peptides.

The present inventors have overcome a number of problems of the prior art.

Summary of the Invention

According to a first aspect of the present invention there is provided a protein interaction system said system comprising a first construct which encodes a first fragment of fluorescent protein, a first bait peptide and a linker portion encoding at least 5 amino acid residues interposed between the first fragment and the bait peptide and a plurality of second constructs encoding a second fragment of fluorescent protein, a prey peptide and a linker portion encoding at least 5 amino acid residues interposed between the second fragment and the prey peptide and on interaction of the bait and a prey peptide the first and second fragments of the fluorescent protein complement each other such that functional association of the first and second fragments promotes fluorescence, wherein at least two of the prey proteins have different amino acid sequences.

Preferably all the prey proteins have different amino acid sequences.
Preferably the first and / or second construct comprises a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the linker of the first and / or second construct is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker of the first and / or second construct is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker of the first and / or second construct is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker of the first and / or second construct encodes up to 60 amino acids.

Where the peptides joined to the linkers are rod like structures and the peptides interact with each other with favourable topology of interaction, ie the peptides interact such that the fragments of

fluorescent protein are brought into close proximity with each other, short linker lengths are sufficient to allow screening for interaction partners. For example, short linkers could be used to screen a library of DNA binding proteins which from previous studies are known to be rod like in structure.

However, linker lengths between 15 to 100 amino acids are advantageous over shorter linker lengths as they allow bulkier peptides being tested for interaction to be conjoined to the first and second fragments of the fluorescent protein without the peptides being tested placing constraints on the functional association of the fluorescent proteins due to stereochemical hindrance. Such longer linkers are also advantageous to study small peptide pairs that have an unfavourable topology of interaction such as is found in an anti-parallel complex (hapto-GFP-N^1->C^1*:binding to :*C^2->N^2-hapto-GFP) i.e. functional association of the interacting peptides causes the fluorescent fragments to be orientated such that they are directed away from each other in space.

Any fluorescent protein may be used in the invention. However, in a preferred embodiment the fragments of fluorescent protein are fragments of green fluorescent protein, mutants or variants thereof.

More preferably the fluorescent protein is the humanised form of a fluorescent protein, e.g.

Enhanced Green Fluorescent Protein (EGFP) or a variant thereof.

Variants include peptides in which individual amino acids are substituted by other amino acids which are closely related as understood in the art, for example, substitution of one hydrophobic residue such as isoleucine, valine, leucine or methionine for another, or the substitution of one polar residue for another, such as arginine for lysine, glutamic for aspartic acid or glutamine for asparagine.

In a humanised nucleotide sequence one or more of the codons in the sequence are altered such that for the amino acid being encoded, the codon used is that which most frequently appears in humans. This is advantageous as the humanised fluorescent protein construct e.g. (EGFP) has maximised expression levels and rate of flurophore formation in mammalian cells. This makes detection of fluorescence, produced by fragments of fluorescent proteins (fluorogenic fragments) which functionally associate with each other, easier to determine.

In a second aspect, there is provided a library of constructs encoding a fragment of fluorescent protein, a peptide and a linker portion of at least 5 amino acids interposed between said fragment and peptide wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such

that on functional association of said fragments fluorescence is enabled wherein the library of constructs encodes a plurality of different peptides.

Each member of the library encodes or provides a different peptide fused to a fragment of fluorescent protein via a linker. The peptides can be small peptides of differing amino acid sequence, for example nonomers, comprising different amino acid compositions or the same overall composition but with the amino acids present in a different order. Alternatively the peptides may be full size proteins obtained from a cDNA library.

Preferably the constructs of the library comprise a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the linker is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker portion is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and

yet more preferably greater than 55 amino acids in length.

Preferably, the linker encodes up to 60 amino acids.

The invention further provides in a third aspect a library of polypeptides, each polypeptide comprising a fragment of fluorescent protein, a peptide and a linker portion of at least 5 amino acid residues interposed between the fragment and the peptide of the polypeptide.

Unless the context demands otherwise, the term peptide, polypeptide and protein are used interchangeably to refer to amino acids in which the amino acid residues are linked by covalent peptide bonds or alternatively (where post-translational processing has removed an internal segment) by covalent di-sulphide bonds, etc. The amino acid chains can be of any length and comprise at least two amino acids, they can include domains of proteins or full-length proteins. Unless otherwise stated the terms, peptide, polypeptide and protein also encompass various modified forms thereof, including but not limited to glycosylated forms, phosphorylated forms etc.

Polypeptides may be made synthetically or recombinantly using techniques which are widely available in the art.

In preferred embodiments, the fragments of fluorescent protein (fluorogenic fragments) are generatable through the introduction of a split point between the amino acids at positions 157 and 158, or (in a second embodiment) between the amino acids at positions 172 and 173 of the humanised form of Green Fluorescent Protein (SEQ ID NO 1).

SEQ ID NO 1 - EGFP (Clontech Inc.) [Genebank Accession number gb:AAB02574 gi 1377912]:

```
1   mvskgeelft gvvpilveld gdvnghkfsv sgegegdaty
41  gkltlkfict tgklpvpwpt lvttltygvq cfsrypdhmk
81  qhdffksamp egyvqertif fkddgnyktr aevkfegdtl
121 vnrielkgid fkedgnilgh kleynynshn vyimadkqkn
161 gikvnfkirh niedgsvqla dhyqqntpig dgpvllpdnh
201 ylstqsalsk dpnekrdhmv llefvtaagi tlgmdelyk
```

The fluorogenic fragments generated by the introduction of a split point between the amino acid residues at positions 157 and 158, or between amino acid residues at positions 172 and 173, result in the production of hapto-EGFP$^{1/157}$ and hapto-EGFP$^{158/239}$, or hapto-EGFP$^{1/172}$ and hapto-EGFP$^{173/239}$, respectively.

Alternative split points are between residues 23/24, 38/39, 50/51, 76/77, 89/90. 102/103, 116/117, 132/133, 142/143, 190/191, 211/212, 214/215 of EGFP.

Thus in preferred embodiments, the fragment comprises a fluorogenic fragment of amino acid residues 1 to 23, 1 to 38, 1 to 50, 1 to 76, 1 to 89, 1 to 102, 1 to 116, 1 to 132, 1 to 142, 1 to

157, 1 to 172, 1 to 190, 1 to 211, 1 to 214, 24 to 239, 39 to 239, 51 to 239, 77 to 239, 90 to 239, 103 to 239, 117 to 239, 133 to 239, 143 to 239, 158 to 239, 173 to 239, 191 to 239, 212 to 239, or 215 to 239 of EGFP.

In one preferred embodiment a library of polypeptides according to a further aspect of the invention is provided wherein each member of the library has a different peptide sequence fused to the fragment of fluorescent protein via the linker region.

A bait peptide is a sequence of two or more amino acids, at least one domain of a protein or a full length protein.

A prey peptide is a sequence of two or more amino acids, at least one domain of a protein or a full length protein.

The term interaction or interacting as used herein means that two entities, for example, distinct peptides, domains of proteins or complete proteins, exhibit sufficient physical affinity to each other so as to bring the two interacting entities physically close to each other. An extreme case of interaction is the formation of a chemical bond that results in continual, stable proximity of the two entities. Interactions that are based solely on physical affinities, although usually more dynamic than chemically bonding interactions, can be equally

effective at co-localising independent entities. Physical affinities include, but are not limited to, for example electrical charge differences, hydrophobicity, hydrogen bonds, van der Waals force, ionic force, covalent linkages, and combinations thereof. The interacting entities may interact transiently or permanently. Interaction may be reversible or irreversible. In any event it is in contrast to and distinguishable from natural random movement of two entities. Examples of interactions include specific interactions between antigen and antibody, ligand and receptor etc.

In a fourth aspect of the invention there is provided a protein interaction monitoring system, said system comprising a first polypeptide comprising a first fragment of fluorescent protein, a bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide and a plurality of second polypeptides comprising a second fragment of fluorescent protein, a prey peptide and a linker portion of at least 5 amino acid residues interposed between the second fragment and the prey peptide and on interaction of the bait and a prey peptide the first and second fragments of the fluorescent protein complement each other such that functional association of the first and second fragments promotes fluorescence, wherein at least two of the prey proteins have different amino acid sequences.

Preferably the linker portion of the first and / or second polypeptide comprises between 15 to 100 amino acid residues.

Preferably the linker of the first and / or second polypeptide is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker of the first and / or second polypeptide is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker portion of the first and / or second polypeptide is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker of the first and / or second polypeptide comprises up to 60 amino acids.

According to a fifth aspect of the present invention there is provided an assay method to determine peptide to peptide interactions comprising the steps of:

providing a first construct, said construct encoding a first fragment of fluorescent

protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a plurality of second constructs said constructs encoding a second complementary fragment of fluorescent protein, a prey peptide and a linker portion of at least 5 amino acids interposed between the second fragment and the prey peptide wherein at least two constructs encode different prey proteins;

expressing both constructs in the same cell; and

detecting fluorescence produced in the cell.

Preferably all the second constructs encode different prey proteins.

Preferably the first and / or second construct comprises a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the first and / or second linker is comprised of substantially hydrophillic amino-acid residues.

More preferably the first and / or second linker encodes multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker of the first and / or second construct is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker of the first and / or second construct encodes up to 60 amino acids.

In an embodiment of the assay the fluorescence detected may be quantitatively determined such that fluorescence produced by different cells or under different conditions can be compared.

In one embodiment of the assay, the second construct is provided as a member of a library of second constructs wherein each member of the library encodes a different prey peptide wherein at least one second construct member of the library is expressed in the same cell as the first construct encoding the bait protein.

The assay can therefore be used to screen an expression library to determine those peptides which bind to a bait peptide.

There is also provided an assay to determine peptide to peptide interactions comprising the steps of:

providing a first polypeptide comprising a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a plurality of second polypeptides comprising a second fragment of fluorescent protein which is complementary to the first fragment of fluorescent protein, a prey peptide and a linker portion of at least 5 amino acids interposed between the second fragment and the prey peptide wherein at least two second polypeptides encode different prey proteins;

mixing the first polypeptide and second polypeptide together; and

detecting whether fluorescence is produced.

Preferably the first and / or second polypeptide linker portion comprises between 15 to 100 amino acid residues.

Preferably the first and / or second polypeptide linker is comprised of substantially hydrophillic amino-acid residues.

More preferably the first and / or second polypeptide linker is comprised of multiples of a

pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the first and / or second polypeptide linker portion is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the first and / or second polypeptide linker comprises up to 60 amino acids.

As detailed above the detected fluorescence can be quantitatively measured.

In a particular example the assay method is performed *in vitro*.

The assay method may further comprise the step of determining the location of the fluorescence in the cell. This is advantageous as it provides details of not only if a protein to protein interaction is occurring, but the location in the cell the interaction is taking place, for example at the membrane, in the cytoplasm, or in the nucleus.

In addition, the assay method may further comprise the step of isolating the bait and / or prey peptide encoded from the cell in which fluorescence has

resulted, for example isolating a cell using a fluorescence activated cell sorting machine then isolating and sequencing the interacting peptides. The sequenced peptides can then be compared with sequences (full length or partial) in a data bank so as to identify or characterise the interacting peptide isolated from the cell.

The sequences of the interacting peptides may alternatively be inferred by cloning selected fluorescent cells and subjecting the cloned cells to PCR amplification and DNA sequencing. These sequences can then be cloned into expression vectors and the protein expressed and purified. The purified protein can be further studied or used for example in research.

In one embodiment, the assay method may further comprise the process of determining the subcellular dynamics of the peptide interactions visualised by fluorescence observations of living cells to enable spatio-temporal studies of peptide interactions throughout all parts of the cell cycle.

In a sixth aspect, which enables spatio-temporal studies, the invention provides an assay which comprises the steps of providing a first construct encoding a polypeptide comprising a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fluorogenic fragment and the first bait peptide:

providing a second construct encoding a polypeptide comprising a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a second prey peptide and a linker portion interposed between the second fluorogenic fragment and the second prey peptide;

causing the expression of both constructs within the same living cell; and

and observing the level of fluorescence produced and its subcellular location in the cell at a range of time points following co-expression of both constructs.

Preferably the first and / or second construct comprises a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the linker of the first and / or second construct is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker of the first and / or second construct is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker of the first and / or second construct is greater than 20 amino acids,

more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker of the first and / or second construct encodes up to 60 amino acids.

In a seventh aspect, there is provided an assay for estimating the maximum possible separation of the fusion termini of the interacting peptides:

providing a first construct encoding a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a second construct encoding a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a prey peptide and a library of linkers of lengths ranging from 5 to 100 amino acids;

expressing both constructs in the same cell following co-transfection of a large population of cells with both constructs;

measuring fluorescence produced in the cell, selection of those cells with higher fluorescence, using either a fluorescence activated cell sorting machine or alternatively by employing laser microdissection; and

clonally amplifying these fluorescent cells, and sequencing the region of a large sample of the constructs encoding the linkers and determining the length of the linkers.

Preferably the linkers of the first and / or second construct are comprised of flexible pentapeptide sequences.

Preferably the pentapeptide is comprised of substantially hydrophillic amino-acid residues. More preferably the pentapeptide is a sequence such as glycyl-glycyl-glycyl-glycyl-serine.

Preferably the number of peptapeptide sequences in the linker is determined by sequencing.

Preferably the linker of the first and / or second construct length is between 10 to 100 amino acids. Alternatively the linker can be between 15 to 100 amino acids in length. In a yet further alternative the linker can be 20 to 100 amino acids in length. As a further alternative the linker can be 30 to 100 amino acids in length.

A distribution of occurrence of linker lengths will be obtained in the fluorescent cells selected, with a sharp cutoff at the lower limit reflecting the minimum linker length capable of spanning the separation of the fusion termini of the interacting peptides and thus allowing productive association of the fluorogenic fragments. A maximum value for this distance may be evaluated in Ångstroms on the basis that each amino acid residue contributes 3.7Å to the length of each linker in an extended backbone conformation.

Further assay methods of the present invention may be used to detect the interactions of three or more agents in a trimeric or higher order complex.

In one example, three fluorescent fragments may provided by introducing two split points as discussed above into the fluorescent protein, each fragment being fused to a peptide. On interaction of the peptides the three or more fluorescent fragments are brought together such that they can functionally associate and generate a fluorescent signal capable of being detected.

In another example one or more of the three fluorescent fragments can be fused to a test agent such as a small ~~molecule, such~~ as a metal ion. In this manner, protein interactions which require the participation of additional test agents, such as small molecules can be detected.

Modulation of the interaction between peptides may be a desirable outcome in the treatment of certain clinical conditions, or as a research tool to study peptide to peptide interactions. For example, modulation of protein to protein interactions may facilitate the task of determining the steps of complex pathways by the provision of means to promote or inhibit a specific interaction, allowing the effects of other proteins to be studied in better detail.

Many protein to protein interactions require the participation of small molecules or peptides. Such a requirement can be determined by simply adding small molecule ligands or the peptides to the components of the assay to determine if these modulate protein to protein interaction as measured by an alteration in fluorescent signal.

Thus in an eighth aspect there is provided an assay for determining whether a candidate agent modulates protein to protein interactions comprising the steps:

providing a first construct encoding a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a second construct encoding a second fragment of fluorescent protein which is

complementary to said first fluorescent fragment, a prey peptide and a linker portion of at least 5 amino-acids interposed between the second fragment and the prey peptide;

providing a putative modulating agent;

expressing both constructs in the same cell; and

measuring fluorescence produced in the cell in the presence and absence of said putative modulating agent

wherein a reduction in fluorescence in the presence of said modulating agent compared to fluorescence in the absence of said candidate modulating agent is indicative of inhibition of complex formation by the modulating agent and an increase in fluorescence is indicative of enhancement of complex formation by the modulating agent.

Preferably the linker of the first and / or second construct comprises a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the linker of the first and / or second construct is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker of the first and / or second construct is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker of the first and / or second construct is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker of the first and / or second construct encodes up to 60 amino acids.

In a ninth aspect there is provided an assay for determining whether a candidate agent modulates protein to protein interactions comprising the steps:

providing a first polypeptide comprising a first fragment of fluorescent protein, a bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a second polypeptide comprising a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a prey peptide and a linker portion

of at least 5 amino-acids interposed between the second fragment and the prey peptide;

providing a putative modulating agent; and

measuring fluorescence produced in the presence and absence of said putative modulating agent

wherein a reduction in fluorescence in the presence of said modulating agent compared to fluorescence in the absence of said candidate modulating agent is indicative of inhibition of complex formation by the modulating agent and an increase in fluorescence is indicative of enhancement of complex formation by the modulating agent.

Preferably the linker of the first and / or second polypeptide comprises between 15 to 100 amino acid residues.

Preferably the linker of the first and / or second polypeptide is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker of the first and / or second polypeptide is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker of the first and / or second polypeptide is greater than 20 amino acids,

more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker of the first and / or second construct polypeptide is up to 60 amino acids.

Thus the above assay can be used to select compounds capable of triggering, stabilising or destablising peptide to peptide interactions.

As will be apparent, the assay of the present invention can be applied in a format appropriate for large scale screening, for example, combinatorial technologies can be employed to construct combinatorial libraries of small molecules or peptides to test as modulating agents.

Preferably, structural information on the peptide to peptide interaction to be modulated is obtained by testing different agents to determine if they are modulating agents.

For example, each of the interacting pair can be expressed and purified and then allowed to interact in suitable *in vitro* conditions. Optionally the interacting peptides can be stabilised by crosslinking or other techniques. The interacting complex can be studied using various biophysical

techniques such as X-ray crystallography, NMR, or mass spectrometry. In addition, information concerning the interaction can be derived through mutagensis experiments for example alanine scanning, or altering the charged amino acids or hydrophobic residues on the exposed surface of the bait or prey peptide being tested.

Based on the structural information obtained, structural relationships between the interacting peptides as well as between the modulating compound and the interacting peptides can be elucidated. Further, the three dimensional structure of the interacting moieties and / or that of the modulating compound can provide information to determine suitable lead compounds able to modulate interaction, which medicinal chemists can use to design analog compounds having similar moieties and structures.

In a tenth aspect, the invention provides compounds obtainable by an assay of the invention, for example small molecules, peptides or nucleic acids which interact with the peptides being tested and modulate the formation of a peptide_complex.

Modulator compounds obtained accordingly to the method of invention may be prepared as a pharmaceutical preparation or composition. Such preparations will comprise the modulating compound and a suitable carrier, diluent or excipient. These preparations may be administered

by a variety of routes, for example, oral, buccal, topical, intramuscular, intravenous, subcutaneous or the like.

According to an eleventh aspect of the present invention there is provided a method of manufacturing a composition or preparation comprising:

performing an assay for determining whether a candidate agent modulates peptide to peptide interactions as described above; and

formulating said agent into a composition.

Also provided are nucleic acid constructs for use in the invention.

Accordingly, in a twelfth aspect, there is provided a nucleic acid construct encoding a fragment of a fluorescent protein, a peptide and a linker portion of at least 15 amino acid residues interposed between said fragment and said peptide, wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled.

Preferably the first and / or second construct comprises a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the linker is comprised of substantially hydrophillic amino-acid residues.

More preferably the linker is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker portion encodes greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

According to a thirteenth aspect of the invention there is provided an expression vector comprising at least one construct encoding a fragment of a fluorescent protein, a peptide and a linker portion of at least 15 amino acid residues interposed between said fragment and said peptide, wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled operably linked to at least one regulatory sequence for the expression of the construct.

The vector can be introduced into the cell using any known techniques such as calcium phosphate

precipitation, lipofection, electroporation and the like.

Where two vectors are provided, and each vector encodes a different construct, for example a bait construct and a prey construct, the vectors can be transfected into the same cell or alternatively into two different cells which are subsequently fused together by cell fusion or other suitable techniques.

In a fourteenth aspect of the invention there is provided a cell transformed with a vector comprising at least one construct encoding a fragment of a fluorescent protein, a peptide and a linker portion of at least 15 amino acid residues interposed between said fragment and said peptide, wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled operably linked to at least one regulatory sequence for the expression of the construct.

Cells which may be transformed include eukaryotic cells, such as yeast, insect, plant, mammalian, primate and human cells. Mammalian cells may be primary cells or transformed cells, including tumour cells. The system is not restricted to intracellular (single cell) interactions. In multicellular organisms amenable to genetic manipulation, a protein-hapto-GFP construct could be released from

one cell or organ and be recognised by another protein-(receptor)-haptoGFP fusion to indicate localisation of filled receptors by the resultant fluorescent signal.

In cell free systems such additional proteins as required for expression may be included, for example, by being provided by expression from suitable recombinant expression vectors.

In addition, there is provided in a fifteenth aspect of the invention a polypeptide encoded by a construct encoding a fragment of a fluorescent protein, a peptide and a linker portion of at least 15 amino acid residues interposed between said fragment and said peptide, wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled.

In a sixteenth aspect of the invention there is provided a library of polypeptides as encoded by constructs according to the fifteenth aspect of the invention.

Preferably the polypeptides of the library comprise a linker portion which encodes between 15 to 100 amino acid residues.

Preferably the polypeptides of the library comprise a linker of substantially hydrophillic amino-acid residues.

More preferably the linker is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the linker portion is greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

Preferably, the linker comprises up to 60 amino acids.

Preferably the first and / or second linker of the vector, or polypeptide can comprises between 15 to 100 amino acid residues.

Preferably the first and / or second linker is comprised of substantially hydrophillic amino-acid residues.

More preferably the first and / or second linker is comprised of multiples of a pentapeptide sequence such as glycyl-glycyl-glycyl-glycyl-serine.

More preferably the first and / or second linker portion encodes or comprises greater than 20 amino acids, more preferably greater than 25 amino acids, more preferably greater than 30 amino acids, more preferably greater than 35 amino acids, even more preferably greater than 40 amino acids, even more preferably greater than 50 amino acids and yet more preferably greater than 55 amino acids in length.

According to a seventeenth aspect of the present invention there is provided a kit comprising at least two constructs according to the twelfth aspect of the invention and means to express the constructs.

The kit may further include test agents, which may enhance or inhibit peptide to peptide interaction.

In another embodiment the kit includes cell lines in which the vector of the thirteenth aspect can be expressed.

Alternatively the kit can comprise at least one polypeptide of the fifteenth aspect of the invention and means for introducing the polypeptide into a cell.

Additionally, ~~the kit can~~ include instructions for using the kit to practise the present invention. The instructions should be in writing in a tangible form or stored in an electronically retrievable form.

Brief description of the figures

The present invention will now be described with reference to the following non-limiting examples and with reference to the figures, wherein:

Figure 1a is a ribbon diagram of EGFP annotated with split point sites;

Figure 1b is an illustration of the split points and the related sequences surrounding these split points of EGFP;

Figure 2 is a representation of a hapto-EGFP with a 26 residue linker between the fluorogenic fragments and the bait and prey proteins respectively;

Figure 3 is a graph of the fluorescence produced by the association of fragments joined to linkers of different lengths, (A) Cells cotransfected with pN157(6)zip and pzip(4)C158 in which a functional leucine zippers mediate the association of haptoEGFP1-157 and haptoEGFP158-238 to generate fluorescence, (B) Negative control cotransfection using pN157(6) and p(4)C158 which lack sequences encoding the leucine zippers and as such fail to generate fluorescence, (D) Cells cotransfected with pN172(6)zip and pzip(4)C173 in which a functional leucine zipper mediated association

of haptoEGFP1-172 and haptoEGFP173-238 occurs to generate fluorescence which is of greater intensity to that observed with the 157/158 split point (E) Negative control cotransfection using pN172(*6*) and p(*4*)C173 which lack sequences encoding the leucine zippers and as such fail to generate fluorescence, (C and F) Confocal images of cotransfected cells from (A) and (D) showing the intracellular localisation of fluorescence. Vero cells were cotransfected with plasmids encoding linkers ranging in length from *4* to *26* amino acids and UV images were collected at 24 hours post-transfection using identical exposure times, (G) pN157(*6*)zip and pzip(*4*)C158 (H) pN157(*16*)zip and pzip(*14*)C158 (I) pN157(*26*)zip and pzip(*24*)C158 (J) pN157(*26*)zip and pzip(*4*)C158 (K) pN157(*6*)zip and pzip(*24*)C158 (L) a negative untransfected control illustrates the background fluorescence level, Italicised figures in brackets indicate the length of the hydrophilic linker;

Figure 4 shows the importance of being able to fuse the interacting peptide to either the N, N′, C or C′ of the fluorescent fragment.

Structural studies of GFP have revealed that the protein exists as a compact cylindrical structure, with eleven beta-sheet strands forming the walls of the cylinder, the N and C termini being at close proximity at the base of the structure. Sections of

alpha-helix form caps on the end of the cylinders and an irregular alpha-helical segment also provides a scaffold for the fluorophore which is located in the geometric center of the cylinder. This folding motif, with beta-sheet outside and helix inside is known as beta-can.

The inventors have shown that fluorescence can be generated following functional association of two separate fragments of GFP molecules (haptoGFPs) when driven by the interactions of a pair of proteins fused both to the new C' and N' termini of each haptoGFP and also to the existing termini.

Functional association of fragments of fluorescent proteins, brought together by the interaction of peptides fused to the fragments, to screen for protein to protein interactions requires that the fragments reliably functionally associate only after interaction of the fused peptides.

Reliable functional association has to date not been obtainable due to the possibility of steric hindrance and steric constraints on the functional association of haptoGFPs when bulky proteins are associated to the GFP fragments.

To overcome problems of steric hindrance, linker regions of at least 15 residues are provided between the peptide being tested for interaction and the associated fluorogenic fragment. This provides the peptide with considerable flexibility relative to

the fluorogenic fragment to bind to another peptide being tested while still enabling the fluorogenic fragments to complement each other and cause detectable fluorescence to be generated.

To prepare GFP fragments, which are capable of functional association, split points were generated at various points along the 239 residue length of the GFP protein, resulting in the generation of new C′ and N′ termini which, in three dimensions, are located at the top and at the base of the beta-can structure.

Split points were introduced based on a structure driven approach between hydrophilic residues. The eleven strands of the beta structure making up the beta-can topology of EGFP are characterised by forming three instances of a tripartite antiparallel sheet motif joined edge to edge around the periphery of the 'can', with the remaining two beta strands completing the cylindrical structure. The most successful split points obtained to date occur in the third tripartite motif between hydrophilic residues allowing adjacent hydrophobic side chains to promote refolding of the haptoGFPs.

As shown in the non exhaustive list of Table 1 a number of split points were identified using the above approach. It would appear that each split point in Table 1 is simply one example of a range of potentially useful split points, the range being shown in parentheses of Table 1.

Table 1

Split point Number	Residue position in EGFP	Possible range
1	23/24	(23-25)
2	38/39	(36-41)
3	50/51	(48-54)
4	76/77	(75-91)
5	89/90	(75-90)
6	102/103	(101-103)
7	116/117	(115-118)
8	132/133	(129-143)
9	142/143	(129-143)
10	157/158	(155-160)
11	172/173	(171-175)
12	190/191	(187-199)
13	211/212	(207-218)
14	214/215	(207-218)

To extend the versatility of the hapto-EGFP method, constructs were created where instead of using C' and N' for the attachment of heterologus proteins, the endogenous termini, N or C together with one of the new N' or C' termini were used. Using this technique the bait and prey peptides can be added such that they are orientated to the associated fluorogenic fragments in the same direction as each other, for example both attached to bottom of the β-can structure of GFP or in the opposite direction, for example the bait peptide is attached to the

bottom of the β-can structure of GFP, while the prey protein is attached to the top of the β-can structure of GFP. As will be understood by those skilled in the art, and as shown in figures 4 A & B, as peptides interact with each other in a particular orientation, then the direction of the linkage of the peptide to the N, N′, C or C′ end of the fluorogenic fragment becomes important in certain circumstances so as to allow the fluorescent protein fragments to functionally interact following interaction of the peptides.

Thus, to minimise interference with the refolding and association of the two hapto-EGFPs during assembly, it would appear that the most versatile split points may occur at the bottom of the β-can. These effects may be minimised by the use of longer linkers to accommodate adverse topology (Figure 4C).

Example 1

As shown in figures 2 and 3, hapto-EGFP with a 26-residue linker between the fluorogenic fragments and the bait and prey proteins respectively were produced without loss of fluorescence. These linkers may be lengthened using overlapping oligonucleotides encoding repeating $(GGGGS)_x$ units. This was achieved by using unique Sac I and BamHI restriction sites present in the core expression vectors pN^{EGFP}(*Sac*)zip and pzip(*Bam*)C^{EGFP}.

To test whether it was possible to obtain fluorescence when endogenous N or C termini and a new N' or C' terminus are used to attach heterologus proteins the fusion (F) and haemagglutinin (H) membrane proteins of measles virus (MV) were used.

Measles virus (MV) infection is mediated by a complex of two viral envelope proteins, haemagglutinin (H) glycoprotein and fusion (F) glycoprotein that bind to each other and then complex with surface receptors to aid the fusion of the virus with the plasma membrane. The H glycoprotein is dimerised in the endoplasmic reticulum and is thought to exist on the cell surface as a tetramer (dimer of dimers). The fusion (F) glycoprotein, is synthesised as an inactive precursor (F_0) which is a highly conserved type I transmembrane glycoprotein of about 60kDa, which is cleaved by furin in the trans-golgi to yield the 41kDa (f_1) and the 18kDa (f_2) disulphide-linked activated F-protein. Infection of the measles virus is dependent on the interaction of the F/H complex with cell surface receptors.

A pair of constructs was generated which encoded the H glycoprotein fused at its N terminus to either the C' terminal residue of hapto-EGFP$^{1-157}$ ($N^{1/157}$) in the first member, or to the natural C terminal residue of the complimentary hapto-EGFP$^{158-239}$ (C158/239) in the second member of the pair. Each construct includes an encoded linker between these two proteins.

Linkers generated using overlapping oligonucleotides which contain *Sfi* IA and *Sfi* IB restriction sites were introduced into $pN^{1/157}(16)zip$ and $pC^{158/239}(14)zip$ constructs. The complete H gene ORF was amplified by polymerase chain reaction (PCR) using primers containing *Sfi* IA and *Sfi* IB restriction sites and the PCR products used to generate $pN^{1/157}(16)MV$-H and $pC^{158/239}(14)MV$-H. A similar pair of constructs employing N and N' hapto-EGFP fusions, pMV-F(16)$N^{1/157}$ and pMV-F(14)$C^{158/238}$ could be generated from existing clones using a similar strategy.

Vero cells (African green monkey kidney-derived cell line) were transiently transfected with $pN^{1/157}(16)MV$-H and pMV-F(14)$C^{158/238}$ constructs, the proteins expressed and phase contrast microscopy used to determine whether the modified glycoproteins retained their fusogenicity.

Real-time observation by ultraviolet and confocal microscopy indicated if fluorescence was generated upon expression of the hapto-EGFP/glycoprotein fusions.

As cells transiently transfected with both F and H expression plasmids form syncytia in the absence of viral replication, the formation of syncytia can be used to assay for successful transfection of both plasmids.

The size of the syncytia was compared with controls to establish whether transfection had occurred. UV

and confocal microscopy were used to examine the fluorescence so as to verify that association between H protein oligomers and F proteins had taken place. Confocal microscopy and image reconstruction were also used to determine the intracellular localisation of H protein oligomers during formation of the fusion complex.

Using the above vectors the intracellular association of F and H proteins and their trafficking from the endoplasmic reticulum (ER) to the plasma membrane was tracked. Further, membrane receptor proteins which interact with the H protein could be identified as could cytoplasmic proteins which interact with known MV receptors and which may therefore initiate downstream signalling events.

Example 2

The above constructs could also be incorporated into a recombinant measles viral genome and the experiments repeated to determine if the above constructs could be used in *in vivo* viral studies,

The type-I F glycoprotein is proposed to form trimers.

In this example two split points could be introduced into the EGFP. The constructs pMV-F(16)$N^{1/157}$, pMV-F(14)$M^{158/190}$ and pMV-F(14)$C^{191/239}$ could be generated.

The method could then be adapted to screen for and identify virus receptors.

This could be tested with MV and applied to the closely related mumps virus (MuV).

Example 3

Fusion of oligonucleotides encoding hapto-EGFP sequences to members of a cDNA library.

Firstly, the sequence encoding the hapto-EGFP may be fused to the 5' end of the library due to the presence of downstream stop codons in the cDNA.

Secondly, constructs are required to be generated for all three reading frames to ensure that one is in the correct reading frame.

Thirdly, the cDNA sequences are required to be obtained from a source which permits directional cloning into restriction sites which are extremely rare in mammalian DNA. Such sequences are to be found in the *Large-Insert cDNA library* (Clontech).

A core panning vector could be engineered from existing plasmids to contain a CMV promoter, an initiation codon and sequences encoding a hapto-EGFP and an intervening linker, an *Sfi* IA site and an *Sfi* IB site, a stop codon and an SV40 polyadenylation signal. Two additional screening vectors could be generated to include one and two extra nucleotides

between the linker and the *Sfi* IA site to correct the reading frame. cDNA fragments, flanked with *Sfi* IA and *Sfi* IB sites obtained from the library could be cloned downstream of the optimised hapto-EGFP linker constructs. The hapto-EGFP library could then be transfected into CHO cells and a mixed population of cells selected using G418 and passaged to confluency. These cells could then be transfected with CD46-haptoEGFP or the equivalent SLAM plasmid.

Where interaction between the peptides being screened occurs, fluorescence is generated.

Any cells which fluoresce can then be isolated by fluorescent laser microdissection and single cell RT-PCR performed to identify mRNA which encodes peptides which interact with the cytoplasmic tails of the receptor molecules.

Although the invention has been particularly shown and described with reference to particular examples, it will be understood by those skilled in the art that various changes in the form and details may be made therein without departing from the scope of the present invention.

Claims

1. A protein interaction system said system comprising a first construct which encodes a first fragment of fluorescent protein, a first bait peptide and a linker portion encoding at least 5 amino acid residues interposed between the first fragment and the bait peptide and a plurality of second constructs encoding a second fragment of fluorescent protein, a prey peptide and a linker portion encoding at least 5 amino acid residues interposed between the second fragment and the prey peptide and on interaction of the bait and a prey peptide the first and second fragments of the fluorescent protein complement each other such that functional association of the first and second fragments promotes fluorescence, wherein at least two of the prey proteins have different amino acid sequences.

2. A protein interaction system as claimed in claim 1 wherein all the prey proteins have different amino acid sequences.

3. A protein interaction system as claimed in claim 1 or claim 2 wherein the first and / or second construct comprises a linker portion which encodes between 15 to 100 amino acid residues.

4. A library of constructs encoding a fragment of fluorescent protein, a peptide and a linker portion of at least 5 amino acids interposed between said fragment and peptide wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled wherein the library of constructs encodes a plurality of different peptides.

5. A library of polypeptides, each polypeptide comprising a fragment of fluorescent protein, a peptide and a linker portion of at least 5 amino acid residues interposed between the fragment and the peptide of the polypeptide wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled wherein the library comprises a plurality of different peptides.

6. A protein interaction monitoring system, said system comprising a first polypeptide comprising a first fragment of fluorescent protein, a bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide

and a plurality of second polypeptides comprising a second fragment of fluorescent protein, a prey peptide and a linker portion of at least 5 amino acid residues interposed between the second fragment and the prey peptide and on interaction of the bait and a prey peptide the first and second fragments of the fluorescent protein complement each other such that functional association of the first and second fragments promotes fluorescence, wherein at least two of the prey proteins have different amino acid sequences.

7. An assay method to determine peptide to peptide interactions comprising the steps of:

providing a first construct, said construct encoding a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a plurality of second constructs said constructs encoding a second complementary fragment of fluorescent protein, a prey peptide and a linker portion of at least 5 amino acids interposed between the second fragment and the prey peptide wherein at least two constructs encode different prey proteins;

expressing both constructs in the same cell; and

detecting fluorescence produced in the cell.

8. An assay method as claimed in claim 7 wherein all the second constructs encode different prey proteins.

9. An assay to determine peptide to peptide interactions comprising the steps of:

providing a first polypeptide comprising a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a plurality of second polypeptides comprising a second fragment of fluorescent protein which is complementary to the first fragment of fluorescent protein, a prey peptide and a linker portion of at least 5 amino acids interposed between the second fragment and the prey peptide wherein at least two second polypeptides encode different prey proteins;

mixing the first polypeptide and second polypeptide together; and

detecting whether fluorescence is produced.

10. An assay which comprises the steps of providing a first construct encoding a polypeptide comprising a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fluorogenic fragment and the first bait peptide:

providing a second construct encoding a polypeptide comprising a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a second prey peptide and a linker portion interposed between the second fluorogenic fragment and the second prey peptide;

causing the expression of both constructs within the same living cell; and

observing the level of fluorescence produced and its subcellular location in the cell at a range of time points following co-expression of both constructs.

11. An assay for estimating the maximum possible separation of the fusion termini of the interacting peptides:

providing a first construct encoding a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5

amino acid residues interposed between the first fragment and the bait peptide;

providing a second construct encoding a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a prey peptide and a library of linkers of lengths ranging from 5 to 100 amino acids;

expressing both constructs in the same cell following co-transfection of a large population of cells with both constructs;

measuring fluorescence produced in the cell, selection of those cells with higher fluorescence, using either a fluorescence activated cell sorting machine or alternatively by employing laser microdissection; and

clonally amplifying these fluorescent cells, and sequencing the region of a large sample of the constructs encoding the linkers and determining the length of the linkers.

12. An assay for determining whether a candidate agent modulates peptide to peptide interactions comprising the steps:

providing a first construct encoding a first fragment of fluorescent protein, a first bait peptide and a linker portion of at least 5

amino acid residues interposed between the first fragment and the bait peptide;

providing a second construct encoding a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a prey peptide and a linker portion of at least 5 amino-acids interposed between the second fragment and the prey peptide;

providing a putative modulating agent;

expressing both constructs in the same cell; and

measuring fluorescence produced in the cell in the presence and absence of said putative modulating agent

wherein a reduction in fluorescence in the presence of said modulating agent compared to fluorescence in the absence of said candidate modulating agent is indicative of inhibition of complex formation by the modulating agent and an increase in fluorescence is indicative of enhancement of complex formation by the modulating agent.

13. An assay for determining whether a candidate agent modulates peptide to peptide interactions comprising the steps:

providing a first polypeptide comprising a first fragment of fluorescent protein, a bait peptide and a linker portion of at least 5 amino acid residues interposed between the first fragment and the bait peptide;

providing a second polypeptide comprising a second fragment of fluorescent protein which is complementary to said first fluorescent fragment, a prey peptide and a linker portion of at least 5 amino-acids interposed between the second fragment and the prey peptide;

providing a putative modulating agent; and

measuring fluorescence produced in the presence and absence of said putative modulating agent

wherein a reduction in fluorescence in the presence of said modulating agent compared to fluorescence in the absence of said candidate modulating agent is indicative of inhibition of complex formation by the modulating agent and an increase in fluorescence is indicative of enhancement of complex formation by the modulating agent.

14. A method of manufacturing a composition or preparation comprising:

performing an assay as claimed in claim 11 or claim 12 for determining whether a candidate

agent modulates peptide to peptide interactions as described above; and

formulating said agent into a composition.

15. A nucleic acid construct encoding a fragment of a fluorescent protein, a peptide and a linker portion of at least 15 amino acid residues interposed between said fragment and said peptide, wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled.

16. An expression vector comprising at least one construct encoding a fragment of a fluorescent protein, a peptide and a linker portion of at least 15 amino acid residues interposed between said fragment and said peptide, wherein said fragment of fluorescent protein is capable of functional association with a complementary fragment of fluorescent protein such that on functional association of said fragments fluorescence is enabled wherein the construct is operably linked to at least one regulatory sequence for the expression of the construct.

17. A cell transformed with a vector as claimed by claim 16.

18. A polypeptide encoded by a construct as claimed in claim 15.

19. A library of polypeptides as encoded by constructs of claim 18 wherein the constructs encode at least two different prey proteins.

20. A kit comprising at least one pair of constructs according to claim 15 wherein the fragment of fluorescent protein of one member of the pair is capable of functional association with a complementary fragment of the other member of the pair and means to express the constructs.

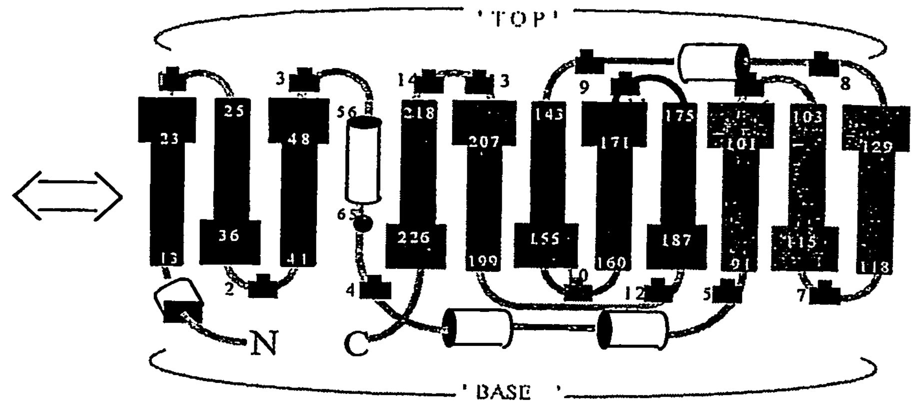
21. A kit as claimed in claim 20 which further comprises test agents, which enhance or inhibit peptide to peptide interaction.

22. A kit as claimed in any one of claims 20 or 21 which further includes a vector of claim 16 and cell lines in which the same can be expressed.

23. A kit as claimed in any one of claims 2 to 22 comprising at least one polypeptide of claim 18 and means for introducing the polypeptide into a cell.

Figure I

A



GFP fold: The ribbon diagram to the left is coloured similarly to the cartoon on the right. β-sheets are indicated by arrows, α-helices by cylinders. Numbers within these symbols refer to sequence positions in EGFP (numbered according to the crystallographic structure - accession: 1emb). In the cartoon, connecting loops are shown by lines. Potential split points are starred and numbered sequentially from the N-terminus, (see **B** below for precise definitions). The fluorophore is represented by a green circle.

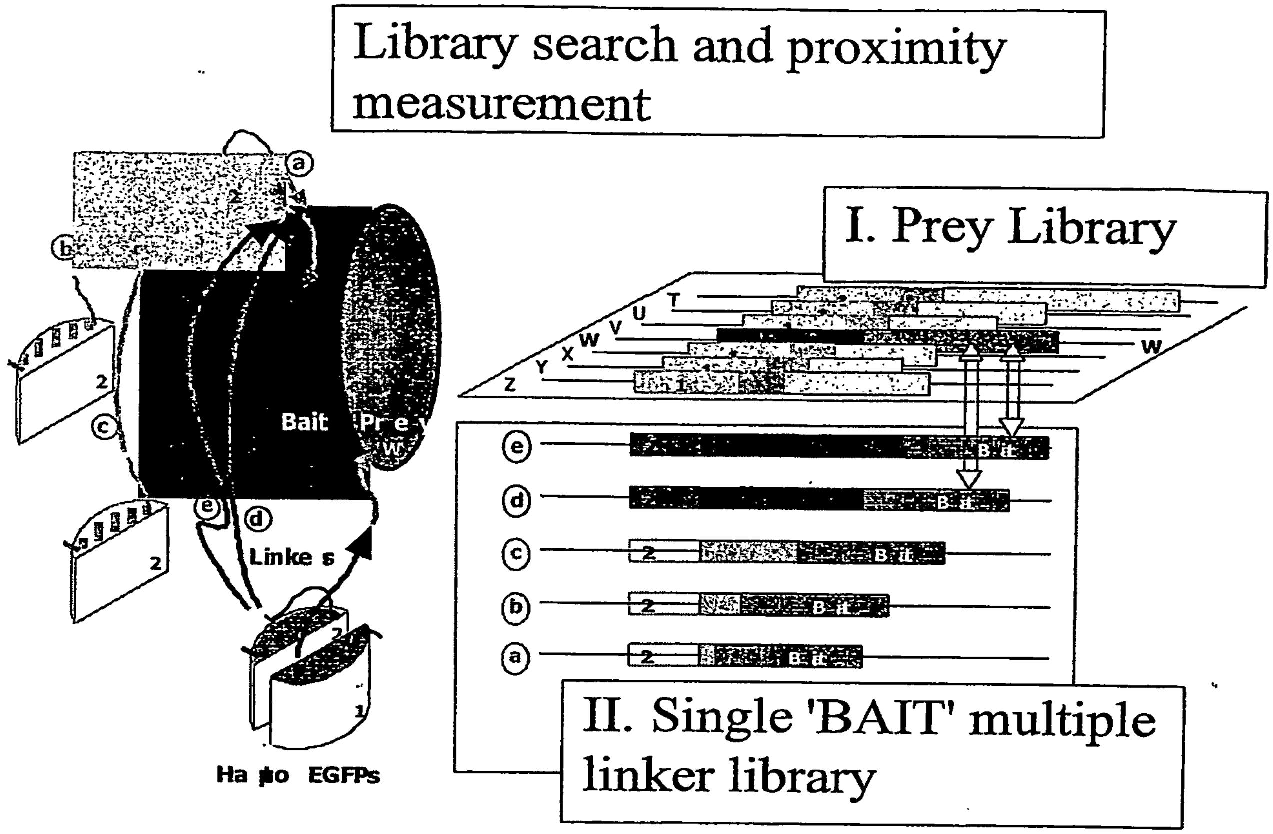
B

Possible split points in EGFP considered for hapto formation.

No.	Position	Top/ Bottom	>----------Sequence---------->
1	23/24	T	...D V N **G**↓**H** K F S...
2	38/39	B	...G D A **T**↓**Y** G K L...
3	50/51	T	...I C T **T**↓**G** K L P...
4	76/77	B	...R Y P **D**↓**H** M K Q...
5	89/90	B	...S A M **P**↓**E** G Y V...
6	102/103	T	...F F K **D**↓**D** G N Y...
7	116/117	B	...K F E **G**↓**D** T L Y...
8	132/133	T	...D F K **E**↓**D** G N I...
9	142/143	T	...H K L **E**↓**Y** N Y N...
10	157/158	B	...A D K **Q**↓**K** N G I...
11	172/173	T	...H N I **E**↓**D** G S V...
12	190/191	B	...P I G **D**↓**G** P V L...
13	211/212	T	...S K D **P**↓**N** E K R...
14	214/215	T	...P N E **K**↓**R** D H M...

Key: Residues at the new, internal C- and N-termini (C' & N') are shown in **bold** with ↓ between them.
Adjacent hydrophobic residues are in *italics*.
The colours in the vertical bar correspond to the structural motifs of the cartoon.

Figure 2

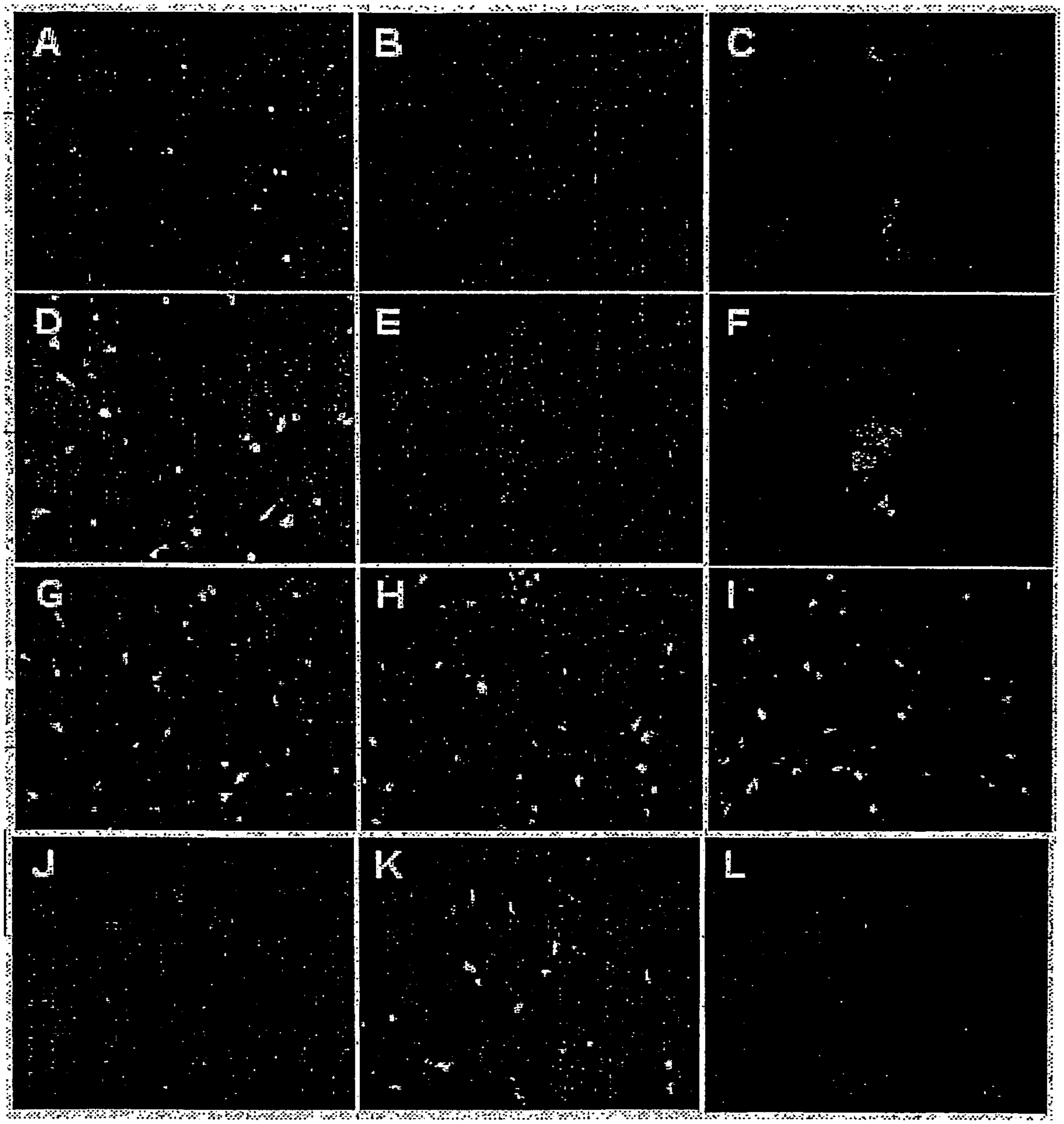


Schematic for protein to protein interaction searches by library interrogation. The two proteins in question are designated 'Bait' and 'W'. Two libraries are generated (I and II), one series of constructs (here designated T….Z, library I, >10,000 members) encodes a hapto-EGFP followed by a DNA sequence encoding a 60 residue linker attached to the 5'-end of a cDNA library, which contains the gene encoding the 'prey', "W" here. The second series of constructs (a…e here, library II, <20 members) encodes the complementary hapto-EGFP followed by a degenerate linker DNA sequence and the 'bait' gene. The individual components of the system are colour coded: blue - 'Bait'/'Prey'; pink - Linker; green - hapto-EGFP. All arrows indicate the direction of the polypeptide backbone (N->C)

A. 'Prey' identification: co-transfection with the 'prey' library (I) and construct '**e**' (long linker - preferably 60 amino acid residues) from the 'bait' library (II) will generate fluorescent cells when the recipient cell receives a vector from library (I) bearing the '**W**' gene (in this case) and a second vector bearing the 'e' bait construct. Clonal expansion of these fluorescent cells allows identification of gene '**W**'.

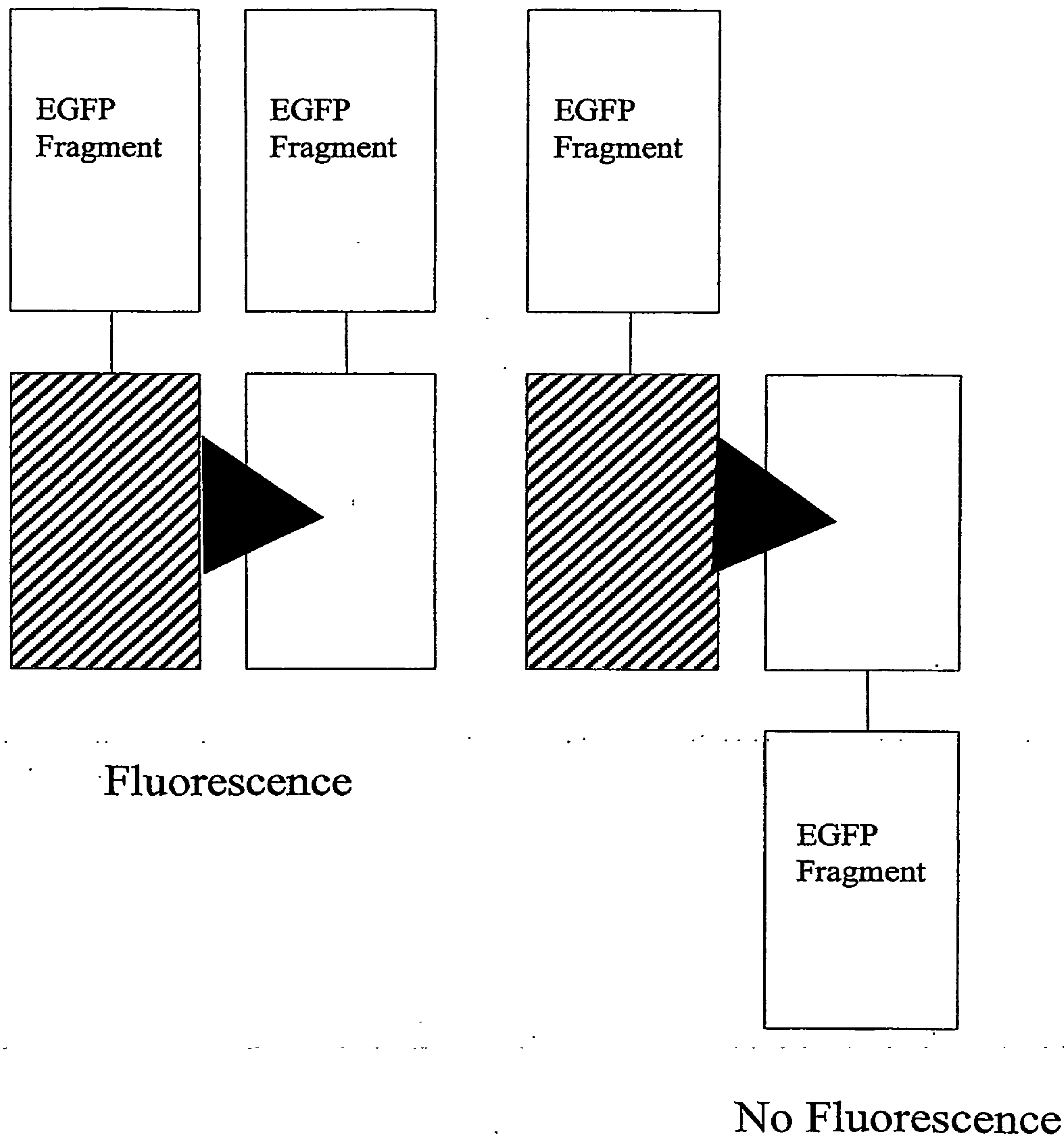
B. Proximity measurement: The clone(s) from step A are co-transfected with the 'bait' library (II). In this case cells showing fluorescence synthesise interacting proteins with a sufficiently long linker to allow productive complementary hapto-GFP interaction. ('**d**' or 'e' in this case), as shown to the left of the diagram. The hollow blue arrows in the right hand part of the diagram are intended to indicate that the interaction of the gene products with these two constructs will generate fluorescence, while other interactions between the product of gene 'W' and the bait protein will not give rise to fluorescent cells due to insufficient length of linker.

Figure 3



Fluorescent images of Vero cells transiently cotransfected with haptoEGFP expression constructs:

Figure 4


EGFP
Fragment
EGFP
Fragment
EGFP
Fragment
Fluorescence
EGFP
Fragment
No Fluorescence